'Charliecloud' simplifies Big Data supercomputing

June 7, 2017

Los Alamos releases elegantly simple high-performance 'convenience bubble'

LOS ALAMOS, N.M., June 7, 2017—At Los Alamos National Laboratory, home to more than 100 supercomputers since the dawn of the computing era, elegance and simplicity of programming are highly valued but not always achieved. In the case of a new product, dubbed "Charliecloud," a crisp 800-line code helps supercomputer users operate in the high-performance world of Big Data without burdening computer center staff with the peculiarities of their particular software needs.

"Charliecloud lets users easily run crazy new things on our supercomputers," said lead developer Reid Priedhorsky of the High Performance Computing Division at Los Alamos. "Los Alamos has lots of supercomputing power, and we do lots of simulations that are well supported here. But we've found that Big Data analysis projects need to use different frameworks, which often have dependencies that differ from what we have already on the supercomputer. So, we've developed a lightweight 'container' approach that lets users package their own user defined software stack in isolation from the host operating system."

To build container images, Charliecloud sits atop the open-source Docker product that users install on their own system to customize the software choices as they wish. Users then import the image to the designated supercomputer and execute their application with the Charliecloud runtime, which is independent of Docker. This maintains a "convenience bubble" of administrative freedom while protecting the security of the larger system. "This is the easiest container solution for both system administrators and users to deal with," said Tim Randles, co-developer of Charliecloud, also of the High Performance Computing Division. "It's not rocket science; it's a matter of putting the pieces together in the right way. Once we did that, a simple and straightforward solution fell right out."

The open-source product is currently being used on two supercomputers at Los Alamos, Woodchuck and Darwin, and at-scale evaluation on dozens of nodes shows the same operational performance as programs running natively on the machines without a container. "Not only is Charliecloud efficient in compute time, it's efficient in human time," said Priedhorsky. "What costs the most money is people thinking and doing. So we developed simple yet functional software that's easy to understand and costs less to maintain."

Charliecloud is very small, only 800 lines of code, and built following two bedrock principles of computing, that of least privilege and the Unix philosophy to "make each

program do one thing well." Competing products range from 4,000 to over 100,000 lines of code. Charliecloud is described in detail in a [technical report online](#).

Los Alamos National Laboratory and supercomputing have a long, entwined history. Los Alamos holds many "firsts," from bringing the first problem to the nation's first computer to building the first machine to break the petaflop barrier. Supercomputers are integral to stockpile stewardship and the national security science mission at Los Alamos.